Empirical study of the anaphoric accessibility space in Spanish dialogues

Patricio Martínez-Barco and Manuel Palomar

Departamento de Lenguajes y Sistemas Informáticos
Universidad de Alicante
Carretera de San Vicente del Raspeig - Alicante - Spain
Tel. +34965903653 Fax. +34965909326
{patricio, mpalomar}@dlsi.ua.es

Abstract

This paper shows an empirical study about the anaphoric accessibility space in Spanish dialogues. According to this study, antecedents of pronominal and adjectival anaphors can almost always (95.9%) be found in the noun phrases set taken from spaces defined using a structure based on adjacency pairs. Furthermore, a proposal of a reliable annotation scheme for Spanish dialogues is presented in order to define this anaphoric accessibility space. Using this annotation scheme, anaphora resolution algorithms can locate the adequate set of anaphor antecedent candidates.

1. Introduction

Anaphora resolution is one of the most active areas of research in Natural Language Processing (NLP). The comprehension of anaphora is an important process in any NLP system, and it is among the toughest problems to solve in Computational Linguistics and NLP.

According to Hirst (1981): *"Anaphora, in discourse, is a device for making an abbreviated reference (containing fewer bits of disambiguating information, rather than being lexically or phonetically shorter) to some entity (or entities) in the expectation that the receiver of the discourse will be able to disabbreviate the reference and, thereby, determine the identity of the entity."*

The reference to an entity is generally called an anaphor (e.g. a pronoun), and the entity to which the anaphor refers is its referent or antecedent. Moreover, it is well-known that anaphora is a mechanism used by speakers in conversation to achieve the common ground. Thus, NLP systems need to both resolve and generate anaphora and they generally resolve it by constructing a set of possible antecedents and then choosing the best one. For this, it is necessary to decide the adequate anaphoric accessibility space, i.e. the space where any anaphora has its candidate set of possible antecedents.

According to Dahlbäck (1991), the efforts made so far towards resolving anaphora can be divided into two basic approaches : Traditional and Discourse-oriented. The traditional approach generally depends on linguistic knowledge. In the discourse-oriented approach, however, the researcher tries to model the complex structure of discourse. Anaphora, accepted as discourse phenomena, is resolved with the help of that complex structure. These works are mostly focused on defining anaphora resolution algorithms, both the traditional approaches (Hobbs, 1986), (Baldwin, 1997), (Mitkov, 1998) and the discourse-oriented ones (Grosz et al., 1995), (Strube and Hahn, 1999).

However, the former do not perform a defined proposal about anaphoric accessibility space, and the latter constraint the space for possible antecedents to the previous utterance. Although, this strategy is adequate for English processing, its application to other languages such as Spanish is not such suitable. For instance, Spanish personal pronouns contain more morphological information than English ones. This makes Spanish speakers to expect larger anaphoric accessibility spaces.

This paper shows that in Spanish dialogues, antecedents of pronominal and adjectival anaphors can almost always be found in the set of noun phrases taken from the anaphoric accessibility space. This space is defined according to an structure based on adjacency pairs (or synchronizing units according to Eckert and Strube (1999a)). Furthermore, a proposal of an annotation scheme for Spanish dialogues is presented, in order to define this anaphoric accessibility space. Moreover, a detailed study of this space and the antecedents we have found in it has been carried out.

Our proposal has been evaluated on the *Corpus InfoTren: Person*, a corpus of Spanish dialogues provided by the BASURDE (1998 2001) Project. These dialogues are conversations between the telephone operator of a railway company and a user of the company.

2. A proposal for an annotation scheme for dialogue structure

For the successful processing and resolution of anaphora in dialogues, we believe that the proper annotation of the dialogue structure is necessary. With such a view, we propose an annotation scheme, for Spanish dialogues, that is based on the work carried out by Gallardo (1996), who applies, to Spanish dialogues, the theories put forward by Sacks et al. (1974) about the taking of speaking turns (conversational). According to these theories, the basic unit of knowledge is the *move* that can inform the listener about an action, request, question, etc. These moves are carried out by means of *utterances*[1]. Therefore, utterances are joined together to become *turns*.

Since our work was done on spoken dialogues that have been written (transcribed), the turn appears annotated in the

[1] An *utterance* in dialogues would be equivalent to a sentence in non-dialogues, although, due to the lack of punctuation marks, utterances are recognized by means of speaker's pauses.

texts and the utterances are delimited by the use of punctuation marks. The reading of a punctuation mark (., ?, !, ...) allows us to recognize the end of an utterance.

As a conclusion, therefore, we propose the following annotation scheme for dialogue structure based on Gallardo (1996):

Turn (T) is identified by a change of speaker in the dialogue; each change of speaker supposes a new speaking turn. On this point, Gallardo makes a distinction between two different kinds of turns:

- An **Intervention Turn (IT)** is one that adds information to the dialogue. Such turns constitute what is called *the primary system of conversation*. Speakers use their interventions to provide information that facilitates the progress of the topic of conversation. Interventions may be **initiatives (IT$_I$)** when they formulate invitations, requirements, offers, reports, etc., or **reactions (IT$_R$)** when they answer or evaluate the previous speaker´s intervention. Finally, they can also be **mixed interventions (IT$_{R/I}$)**, meaning a reaction that begins as a response to the previous speaker's intervention, and ends as an introduction of new information.
- A **Continuing Turn (CT)** represents an empty turn, which is quite typical of a listener whose aim is the formal reinforcement and ratification of the cast of conversational roles. Such interventions lack information.

Adjacency Pair or **Exchange (AP)** is a sequence of turns T headed by an initiation intervention turn (IT$_I$) and ended by a reaction intervention turn (IT$_R$). One form of anaphora which appears to be very common in dialogues is the reference within an adjacency pair (Fox, 1987).

Topic (TOPIC) is the main entity in the dialogue. According to Rocha (1998) four features are taken into account in the selection of the best candidate for discourse topic: frequency, even distribution, position of first token, and semantic adequacy. The topic must be a lexical item which is frequently referred to.

According to the above-mentioned structure, the following set of tags is considered necessary for dialogue structure annotation: IT$_I$, IT$_R$, CT, AP and TOPIC. AP and TOPIC tags will be used to define the anaphoric accessibility space and the remaining will be used to obtain the adjacency pairs. The IT$_{R/I}$ tag standing for mixed interventions is not considered because mixed interventions can be split into two different interventions: IT$_R$ and IT$_I$. This task will be done in the annotation phase.

For this experiment, the corpus has been manually annotated. However, nowadays there are some works performing an automatic adjacency pair tagging, such as the BASURDE (1998 2001) Project. On the other hand, there are other works performing automatic topic tagging (e.g. Reynar (1999)) or automatic topic extraction (e.g. the method for anaphora resolution shown in Martínez-Barco et al. (1999)).

An example of an annotated dialogue with such tags is presented in figure 1. It should be pointed out that the tag (OP) indicates the turn of the operator of a railway company, and the tag (US) indicates the user's turn. The transcribed dialogue provides these tags.

The annotation of conversational dialogues is carried out, as shown above, and the evaluation of the proposed anaphoric accessibility space accomplished. An important aspect of this type of annotation is the training phase, which assures the reliability of the annotation.

The annotation phase is accomplished in the following way: a) two annotators are selected, b) an agreement[2] is reached between the two annotators with regard to the annotation scheme using 5 dialogues (training corpus), c) the annotation is then carried out by both annotators in parallel over the remaining 35 dialogues (test corpus) and, d) finally, a reliability test is done on the annotation (see Carletta et al. (1997)). The reliability test uses the *kappa* statistic that measures the affinity between the annotations of the two annotators by making judgements about categories. See Siegel and Castellan (1988) for *kappa* statistic (*k*) computing.

Because of turns are marked during the transcription phase, all the annotator must do in relation to the adjacency pair is to classify turns according to the above classification, and then to relate each initiative intervention turn IT_I to its reaction intervention turn IT_R. As a result, the adjacency pair is defined. Thus, this task was limited just to a classification task that is easily measured using the *kappa* statistic.

Another task is the topic definition. According to the corpus structure, this task is trivial because the corpus is organized into short dialogues, and each dialogue has only one main topic or theme. This topic is introduced clearly by means of some user's intervention at the beginning of the dialogue. Consequently, we have not detected discrepancies between both annotators with regard to the topic definition, and because of this, this task was not measured using the *kappa* statistic.

According to Carletta, a *k* measurement such as $0.68 < k < 0.8$ allows us to make encouraging conclusions, and $k > 0.8$ means total reliability between the results of both annotators.

Once both annotators have carried out the annotation, the reliability test of the annotation has been run, with a *kappa* measurement of $k = 0.91$. We therefore consider the annotation obtained for the evaluation to be totally reliable.

In those cases where some discrepancy between the annotators was found, the following criteria was applied: each dialogue has a main annotator whose criteria with regard to the annotation is considered definitive although there were discrepancies between both accounts. In order to guarantee the results, each annotator was the main annotator in only 50% of the dialogues.

As this annotation would be processed by some

[2]This agreement is about what every tag means to every annotator when it is applied to the corpus

TOPIC			tren *(train)*
AP1		IT_I (OP)	información de Renfe, buenos días *(Renfe information, good morning)*
		IT_R (US)	hola, buenos días *(hello, good morning)*
		CT (OP)	hola *(hello)*
AP2		IT_I (US)	me podéis decir algún tren que salga mañana por la tarde para ir a Monzón *(could you tell me about any train that leaves tomorrow evening for Monzon)*
		IT_R(OP)	si, vamos, mira hay un talgo a las tres y media de la tarde *(let me see, there is a talgo at half past three)*
AP3		IT_I (US)	sí tiene que ser más tarde *(it has to be later)*
		IT_R (OP)	más tarde. Hay un intercity a las cinco y media, un expreso a las seis y media *(later. There is an intercity at half past five, an expreso at half past six)*
AP4		IT_I (US)	el de las seis y media ¿llega a Monzón? *(the half past six one, does it go to Monzon?)*
	AP5[a]	IT_I (OP)	a ver. El de las seis y media me ha preguntado ¿verdad? *(let me see. You have asked about the half past six one, haven't you?)*
		IT_R (US)	si *(yes)*
		IT_R (OP)	a las nueve y veinticinco *(at twenty-five past nine)*
AP6		IT_I (US)	a las nueve y veinticinco está en Monzón *(at twenty-five past nine it is in Monzon)*
		IT_R (OP)	si *(yes)*
		CT (US)	vale, pues ya está. Esto ya es suficiente. *(ok, that's all. That's enough.)*
		CT (OP)	hum, hum (simultáneo)
AP7		IT_I (US)	gracias, ¿eh? *(thank you, ok?)*
		IT_R (OP)	muy bien a usted. Hasta luego *(thanks. Bye)*

[a] This adjacency pair is included in AP4

Figure 1: An example of an annotated dialogue from *Corpus InfoTren: Person*

anaphora resolution system, we propose an SGML tagging format such as the one that can be seen in figure 2.

The SGML markup will have the following form:

```
<ELEMENT-NAME ATTR-NAME="VALUE" ...>
text-string
</ELEMENT-NAME>
```

Thus, the following notation is provided in each case:

- Topic:

```
<TOPIC>
topic-entity
</TOPIC>
```

- Adjacency pairs:

```
<AP ID="number">
Adjacency-pair
</AP>
```

ID contains an identification number for arranging the adjacency pairs in sequential order.

- Intervention turns:

```
<IT TYPE="R|I" SPEAKER="speaker">
Intervention-turn
</IT>
```

```
<TOPIC>                              tren
                                     (train)
</TOPIC>
                                     ...
<AP ID="4">
<IT TYPE="I" SPEAKER="US">           el de las seis y media ¿llega a Monzón?
                                     (the half past six one, does it go to Monzon?)
</IT>
<AP ID="5">
<IT TYPE="I" SPEAKER="OP">           a ver. El de las seis y media me ha preguntado ¿verdad?
                                     (let me see. You have asked about the half past six one, haven't you? )
</IT>
<IT TYPE="R" SPEAKER="US">           si
                                     (yes)
</IT>
</AP>
<IT TYPE="R" SPEAKER="OP">           a las nueve y veinticinco
                                     (at twenty-five past nine)
</IT>
</AP>
                                     ...
```

Figure 2: SGML annotation example

TYPE may be "R" or "I" (Reaction or Initiative), and SPEAKER is the mark for the participant that is speaking this turn.

- Continuing turns:

```
<CT SPEAKER="speaker">
Continuing-turn
</CT>
```

3. Accessibility space proposal

Based on the above-mentioned annotation, an anaphoric accessibility space is proposed in order to solve anaphors generated by Spanish personal pronouns, demonstrative pronouns and adjectival anaphors [3].

According to Fox (1987) the first mention of a referent in a sequence is done with a full noun phrase. After that, by using an anaphor the speaker displays an understanding that sequence has not been closed down. Then, we consider that two different sequences generate mostly of the anaphors to be found in dialogues: the adjacency pair and the topic scope. The former generates references to any local noun phrase, and the later generates references to the main topic of the dialogue.

Based on this, we propose the anaphoric accessibility space as the set of noun phrases taken from:

- the same adjacency pair as the anaphor, plus
- the previous adjacency pair to the anaphor, plus
- another adjacency pair including the anaphor adjacency pair, plus
- the noun phrase representing the main topic of the dialogue.

[3] the Spanish adjectival anaphor is a kind of English one-anaphora where the word *one* is omitted. For instance, *el de las seis y media (the half past six one)*.

4. Empirical study

In order to carry out the evaluation of the anaphoric accessibility space, the global process shown in figure 3 was performed.

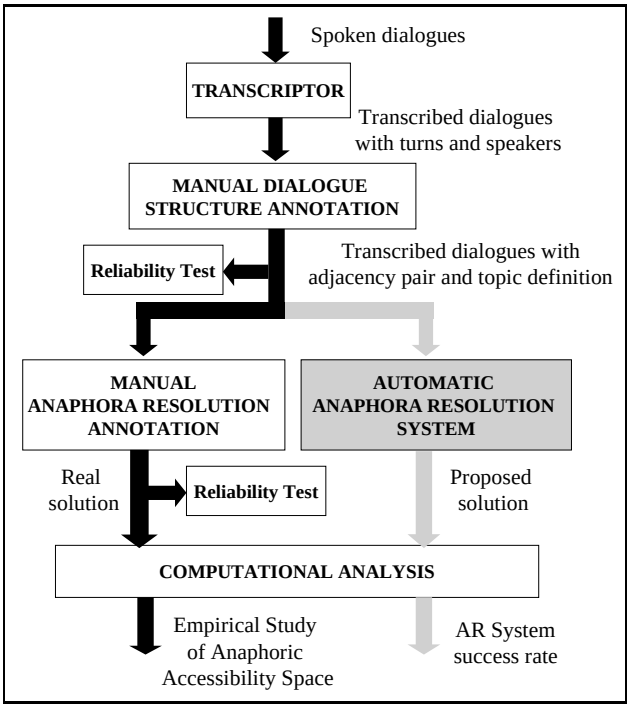

Figure 3: Global process

In this experiment, 40 transcribed spoken dialogues were selected from the 200 afforded us by the *Basurde* project. The transcriptor used in the *Basurde* project provides written dialogues with turn and speaker marks.

	Same AP[a]	Previous AP[b]	Included AP[c]	TOPIC[d]	Others[e]
Pronominal	60.6%	24.6%	8.2%	4.9%	1.7%
Adjectival	44.7%	28.9%	5.2%	13.4%	7.8%
Total Results	Anaphoric accessibility space proposal: 95.9%				4.1%

[a]The antecedent is found in the same Adjacency Pair as the anaphor one
[b]The antecedent is found in the previous Adjacency Pair to the anaphor one
[c]The antecedent is found in another adjacency pair including the anaphor adjacency pair
[d]The antecedent is found in the main Topic of dialogue
[e]The antecedent is found in other sources

Table 1: Empirical results

Afterwards, these selected dialogues were manually annotated according to the proposed annotation scheme. From the 40 dialogues, 5 were randomly selected for the annotators' training and the remaining 35 were reserved in order to carry out the final evaluation. Then, the reliability test of this annotation was performed in order to guarantee the final results.

Following this, a manual annotation of the anaphor solutions was performed over pronominal and adjectival anaphors in the corpus. This annotation relates each anaphor to the correct antecedent. Again, in order to guarantee the results, this annotation was performed by two annotators in parallel, and a reliability test of the annotation was carried out. In this way, the annotation was considered a classification task consisting in defining the adequate solution from the candidate list (we estimated an average of 6.5 possible candidates per anaphora after applying restrictions). Once the reliability test over the manual anaphora resolution annotation was run, a *kappa* measurement of $k = 0.87$ was achieved.

After that, a study of each pronominal and adjectival anaphora was developed to obtain the antecedent location, as shown in table 1. This study was made applying a computational analyzer that obtains information about an automatic anaphora resolution system[4]. As a result, the analyzer compares the output of this AR system with the real solution in the manual annotation and provides several statistics about it. One of these statistics is the study presented in this paper[5].

According to this study, 95.9% of the antecedents were located in the proposed anaphoric accessibility space. Remaining antecedents (4.1%) were estimated to be located in subtopics of the dialogues. In order to incorporate these antecedents to the anaphoric accessibility space, a basic strategy based on the use of the full space (i.e. all the noun phrases from the beginning of the dialogue to the anaphor) could be proposed. As shown in table 2, our proposal of anaphoric accessibility space works with an average of 10.5 antecedents per anaphor (before applying restrictions) instead of 35 antecedents per anaphor that could be obtained if we consider the full space. That means a decreasing of 70%. Evaluating the advantages and the disadvantages, considering the full space implies a) great computational efforts and b) 70% more possibilities to obtain an incorrect response in the anaphora resolution algorithm that uses this anaphoric accessibility space. Notice that our experiments had been performed over a collection of short dialogues (around 332 words per dialogue). This difference will increase in longer dialogues.

[4]This anaphora resolution system uses an algorithm based on the proposed anaphoric accessibility space (see Martínez-Barco and Palomar (2000)).

[5]Notice that the study about anaphoric accessibility space was not developed using the AR system proposal, but the manual annotation of anaphors, (i.e. real solutions).

	Full space	AAS proposal
Total antecedents	3245	1025
Antecedents per anaphor	35	10.5
Reduction	70%	

Table 2: Anaphoric accessibility space vs full text

5. Conclusions

This paper shows that in a corpus of Spanish dialogues, the antecedent of pronominal and adjectival anaphors can almost always be found in the set of noun phrases taken from the same adjacency pair as the anaphor, the previous adjacency pair, any containing adjacency pair, plus a noun phrase representing the main topic of the dialogue when the anaphor occurs.

Furthermore, an annotation scheme of dialogue structure for Spanish has been presented, allowing us to define the adequate anaphoric accessibility space. Starting with the study performed over a dialogue corpus, it has been shown that this proposed space allows us to locate 95.9% of anaphoric antecedents. We consider that anaphora resolution in Spanish dialogues needs to have a dialogue structure and define the adequate space that improves this resolution.

In this work, we only deal with individual anaphora, i.e. anaphors whose antecedents are noun phrases. There are several studies about deictic anaphora, that is, anaphors having abstract antecedents, showing the importance of this kind of anaphora in dialogues (see Eckert and Strube (1999b)). Thus, a full study of spaces for deictic anaphora and other kinds of anaphora (surface-count anaphora, definite descriptions, one-anaphora, etc.) must be performed.

6. Acknowledgments

The authors wish to thank N. Prieto, F. Pla and A. Molina (Universitat Politecnica de Valencia) for having

contributed their tagger and E. Segarra (Universitat Politecnica de Valencia) for affording us the *Corpus InfoTren: Person* from the BASURDE Project. We are also grateful to several anonymous reviewers of *Gotalog'2000* for their helpful comments of earlier drafts of the paper. And finally thanks also to Rafael Muñoz and Maximiliano Saiz Noeda who discussed several issues about the final paper.

This paper has been supported by the Comisión Interministerial de Ciencia y Tecnología (CICYT) with project number TIC97-0671-C02-01/02.

7. References

B. Baldwin. 1997. CogNIAC: high precision coreference with limited knowledge and linguistic resources. In *Proceedings of ACL/EACL workshop on Operational factors in practical, robust anaphora resolution*, Madrid (Spain), July.

Proyecto BASURDE. 1998–2001. *Spontaneus-Speech Dialogue System in Limited Domains*. CICYT (TIC98-423-C06). http://gps-tsc.upc.es/veu/basurde/Home.htm.

J. Carletta, A. Isard, S. Isard, J.C. Kowtko, G. Doherty-Sneddon, and A.H. Anderson. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics*, 23(1):13–32.

N. Dahlbäck. 1991. *Representations of Discourse-Cognitive and Computational Aspects*. Ph.D. thesis, Department of Computer and Information Science, Linköping University, Linköping, Sweden.

M. Eckert and M. Strube. 1999a. Dialogue Acts, Synchronising Units and Anaphora Resolution. In *Proceedings of Amsterdam Workshop on the Semantics and Pragmatics of Dialogue (AMSTELOGUE'99)*, University of Amsterdam, Holland, May.

M. Eckert and M. Strube. 1999b. Resolving Discourse Deitic Anaphora in Dialogues. In *Proceedings of 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL'99)*, Bergen, Norway.

B. Fox. 1987. *Discourse Structure and Anaphora*. Written and conversational English. Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.

B. Gallardo. 1996. *Análisis conversacional y pragmática del receptor*. Colección Sinapsis. Ediciones Episteme, S.L., Valencia.

B. Grosz, A. Joshi, and S. Weinstein. 1995. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

G. Hirst. 1981. *Anaphora in Natural Language Understanding*. Springer-Verlag, Berlin.

J. Hobbs. 1986. Resolving pronoun references. In B. Grosz B.L. Webber and K. Jones, editors, *Readings in Natural Language Processing*. Morgan Kaufmann, Palo Alto, CA.

P. Martínez-Barco and M. Palomar. 2000. Dialogue structure influence over anaphora resolution. In O. Cairo, L.E. Sucar, and F.J. Cantu, editors, *MICAI 2000: Advances in Artificial Intelligence*, volume 1793 of *Lecture Notes in Artificial Intelligence*, Acapulco, México, April. Springer-Verlag.

P. Martínez-Barco, R. Mu ̄noz, S. Azzam, M. Palomar, and A. Ferrández. 1999. Evaluation of pronoun resolution algorithm for Spanish dialogues. In *Proceedings of the Venezia per il Trattamento Automatico delle Lingue (VEXTAL'99)*, pages 325–332, Venice (Italy), November.

R. Mitkov. 1998. Robust pronoun resolution with limited knowledge. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98)*, Montreal (Canada), August.

Jeffrey C. Reynar. 1999. Statistical Models for Topic Segmentation. In *Proceedings of 37th Annual Meeting of the Association for Computational Linguistics*, pages 357–364, Maryland, USA, June.

M. Rocha. 1998. *A corpus-based study of anaphora in dialogues in English and Portuguese*. Ph.D. thesis, University of Sussex, Sussex. UK.

H. Sacks, E. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn taking for conversation. *Language*, 50(4):696–735.

S. Siegel and J. Castellan. 1988. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, 2nd edition.

M. Strube and U. Hahn. 1999. Functional Centering - Grounding Referential Coherence in Information Structure. *Computational Linguistics*, 25(5):309–344.